AFRL-AFOSR-VA-TR-2017-0074

Synergic investigation of the self-assembly structure and mechanism of retroviral capsid proteins by solid state NMR, transmission electron microscopy and multiscale simulation

Bo Chen
UNIVERSITY OF CENTRAL FLORIDA
4000 CNTRL FLORIDA BLVD
ORLANDO, FL 32816

03/29/2017
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTB2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
29-03-2017

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
15 Mar 2013 to 14 Sep 2016

4. TITLE AND SUBTITLE
Synergic investigation of the self-assembly structure and mechanism of retroviral capsid proteins by solid state NMR, transmission electron microscopy and multiscale simulation

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-13-1-0150

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Bo Chen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF CENTRAL FLORIDA
4000 CNTRL FLORIDA BLVD
ORLANDO, FL 32816 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTB2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2017-0074

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION A: Distribution approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Rous Sarcoma Virus (RSV) is the ideal platform to study retrovirus. The RSV capsid enclosing the viral genome materials is assembled from ~ 1500 copies of the 237-residue RSV capsid protein (CA). In vitro, tubular assembly can be obtained with the CA with similar underlying structural properties as the authentic RSV capsid. Due to strong polymorphism, RSV CA assemblies are challenging for structural characterization by techniques such as X-ray diffraction or cryo Electron Microscopy (cryo-EM).

15. SUBJECT TERMS
nmr, self-assembly, retro-virus

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			ROACH, WILLIAM
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 703-588-8302

Final report: Synergic investigation of the self-assembly structure and mechanism of retroviral capsid proteins by solid state NMR, transmission electron microscopy and multiscale simulation

Bo Chen

Department of Physics, University of Central Florida, Orlando, FL bo.chen@ucf.edu

Report period: 15-03-2016 to 14-09-2016
Contract number: FA9550-13-1-0150; 2013 YIP award

Cover page

Dear Program manager and AFOSR,

First I would like to express my most sincere acknowledgement of your support in the past three years, which made all this wonderful scientific results come to life possible.

In this final report, we summarize the achievements in the last half year and how we made it in some details. We overachieved 100% of the goal on characterization of the RSV CA tubular assembly, established the first atomic resolution structural model of the tube by collaborating with cryo-EM experts at University of Auckland. We sequentially assigned half of the residues of RSV CA in the spherical assembly.

The following sections are listed in the report:

A. Abstract
B. Objectives
C. Findings
D. Supported Personnel
E. Collaborations
F. Publications
G. Interactions/Transitions
H. References.

If you have any questions, please do not hesitate to contact me.

Thanks!

Yours, sincerely

Bo Chen
Assistant Professor
Department of Physics
University of Central Florida
Bo.chen@ucf.edu
407-823-4494

Final accomplishments Abstract: The Rous Sarcoma Virus (RSV) is the ideal platform to study retrovirus. The RSV capsid enclosing the viral genome materials is assembled from ~ 1500 copies of the 237-residue RSV capsid protein (CA). In vitro, tubular assembly can be obtained with the CA with similar underlying structural properties as the authentic RSV capsid. Due to strong polymorphism, RSV CA assemblies are challenging for structural characterization by techniques such as X-ray diffraction or cryo Electron Microscopy (cryo-EM).

During the past six months, we developed a novel method to exploit well-resolved NCACX spectra of the RSV CA tubular assembly to assist the resonance assignment of NCOCX spectra. Applying this method, we assigned 234 residues out of the 237 residues RSV CA. In addition, we developed collaboration with cryo-EM specialist at University of Auckland. Combining with constraints from cryo-EM, we established an atomic resolution model of the tubular assembly by molecular dynamics flexible fitting. Our model shows that significant structural rearrangements take place at flexible loops and the 3_{10} helix regions, while the rest of the protein retains its structure upon assembly. The analyses of our model suggests the assembly polymorphism is attributed to the disorder of the trimer interface between C-terminal domains. In addition, the different contact angles between helices at assembly interfaces of tubular and planar assemblies for HIV and RSV CA, which suggests the two system undergo different assembly pathways.

In addition, we performed partial sequential assignments (110 out of 237 residues) of the spherical RSV CA assembly. Additional spectra of the spherical RSV CA assembly will be acquired to complete the assignments

B. Objectives:
1. Experimentally:
(1a) Determine the secondary structure and dynamics of capsid protein in distinct assemblies (*tubular assembly 100% complete*, spherical assembly undergoing),
(1b) Characterize the subunit contacts in different assemblies(*tubular assembly 100% complete*, spherical assembly undergoing).
.
2. Theoretical simulations:
2a) Establish a highly efficient framework with full 3D capability and investigate the correlation between structure and the assembly mechanism at the helix-level (*100% complete*).
2b) Construct a hybrid model to incorporate residue-specific interactions into our CG model to understand the causes of the structural variations (*100% complete*).

C. Findings:

(1). Nearly complete assignments of the RSV CA protein in its tubular assembly

Although the sequential assignments were achieved before Dec, 2016, we took additional time during the past six months to digest the technical insights we achieved in the process, and developed the strategy that can be generalized for ssNMR characterize other protein assemblies with congested spectra.

It is a universal challenge for ssNMR studies of large proteins, to determined accurately the residue-specific assignments (RSA) in crowded spectra. Normally, owing to larger Ca chemical shift (CS) dispersion (~ 25-30 ppm), resonances in NCACX spectra are better resolved than those in NCOCX spectra, as shown in Figure 1. Here we take spectra of the tubular assembly of uniform ^{13}C, ^{15}N labeled RSV CA(U-CA), and 1,3-13C and 2-13C glycerol labeled RSV CA(1,3-G and 2-G). As Figure 1 shows, NCACX spectra are well resolved, but NCOCX spectra remain highly congested even with sparsely labeled protein by glycerol.

To solve this issue, we adopted a novel method that exploits the intrinsic correlation between NCACX and NCOCX spectra to enhance the efficiency and accuracy of RSA. Although all carbon sites in the same residue are correlated with different amide nitrogen in NCACX and NCOCX spectra, they exhibit identical carbon resonances in both. Therefore, the well-resolved NCACX spectra can be used to guide the RSA of congested NCOCX. The implementation of this strategy is shown in Figure 2. As shown in Figure 2A, a large number resonances are present at CO=176.4 ppm (vertical axis) in the extracted ^{15}N=119.1 pm plane of NCOCX spectra. Among all Ca resonances (signals higher than 50 ppm) at this CO frequency in this ^{15}N plane, assume we can assign the residue type for the Ca signal at 61 ppm. What remains to be determined is which of the side-chain carbons (signals lower than ~ 45 ppm) arise from the same residue. Without additional constraints, the problem is intractable as there are too many candidates due to congested signals. However, the intra-residue correlation between all carbons is recorded in NCACX spectra at that Ca frequency plane, albeit they are correlated with a different amide nitrogen than that in NCOCX spectra. Therefore, correct side-chain signals for this 61 ppm Ca resonance at CO=176.4 ppm in ^{15}N=119.1 pm plane of the NCOCX spectra can be identified unambiguously by inspecting the Ca=61 ppm plane of the NCACX spectra, which is well-resolved, shown in Figure 2B. By comparing the side-chain resonances at CO 176.4 ppm slice in the ^{15}N=119.1 plane of the NCOCX spectra and those in the ^{13}C=61 ppm plane in NCACX spectra that exhibit a correlation with the CO resonance of 176.4 ppm, it is clear that only resonances at 44.5, 30.4, and 27.8 ppm along the ^{15}N=115.8 ppm slice meet both criteria, as shown in Figure 2B. It suggests that these resonances arise from the same residue with the Ca signal at 61 ppm and carbonyl signal at 176.4 ppm. Further inspection of the corresponding ^{15}N=115.8 ppm plane of NCACX spectra confirms this intra-residue correlation, as shown in Figure 2C. They are assigned as an R according to the characteristic resonances of amino acids. Therefore, this approach decomposes the congested signals in NCOCX into dispersed resonances from individual residues at respective Ca planes of the well-resolved NCACX spectra for reliable RSA. With this approach, the accuracy and certainty of RSA were greatly improved, and the sequentially assigned residues were increased to 128, in comparison to the 82 sequential assigned residues without this strategy (Eventually 234 residues were assigned with additional samples prepared with RSV CA of ^{13}C, ^{15}N labeling of all R and L residues). More generally, this strategy can be applied to RSA of other proteins with congested 3D spectra.

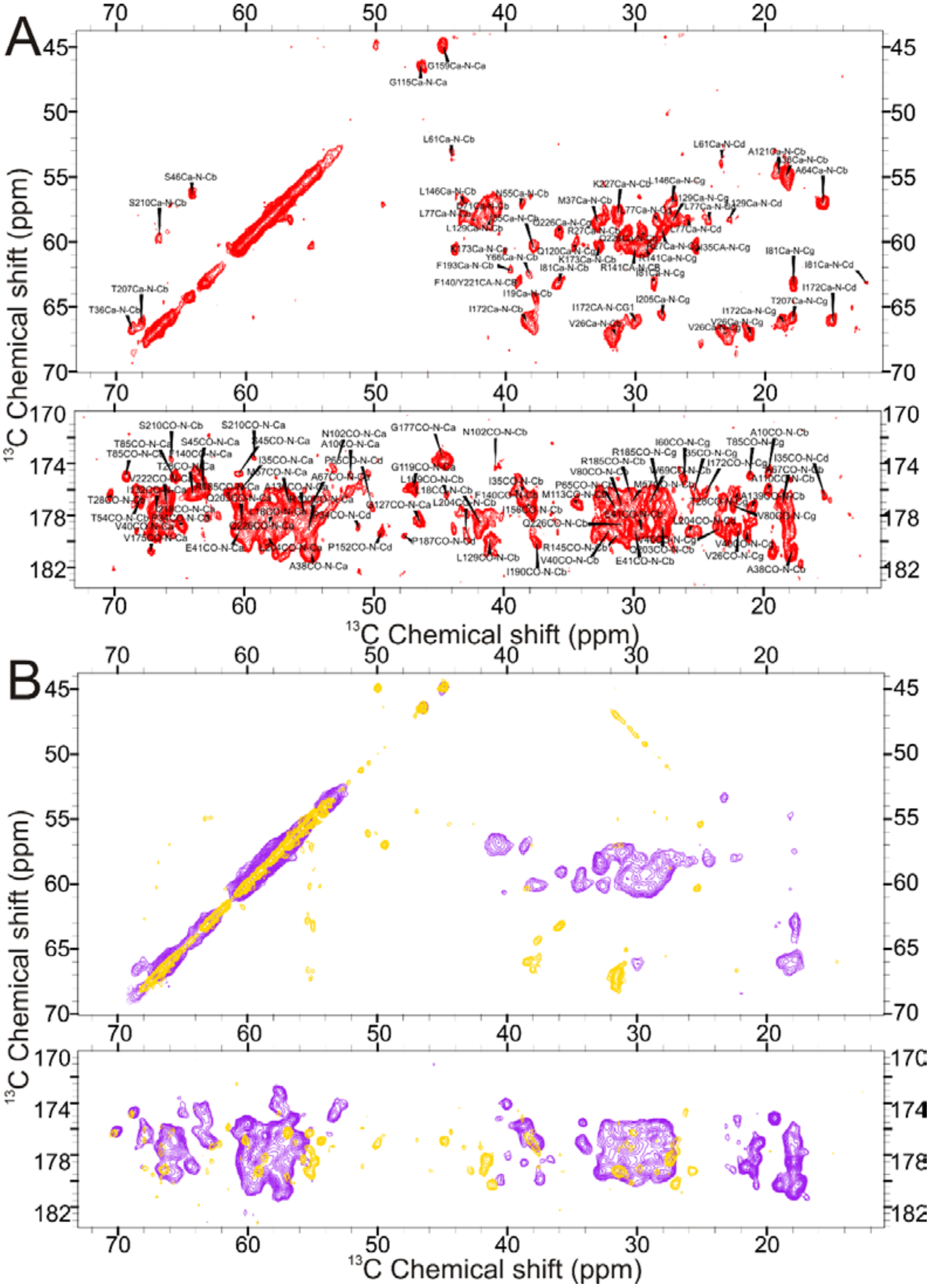


Figure 1. Aliphatic regions of the 2D planes at ^{15}N = 118.95 ppm extracted from 3D NCACX (top panel) and 3D NCOCX spectra (bottom panel) for U-CA sample in (A) and 1,3G-CA (purple) and 2G-CA (yellow) samples in (B).

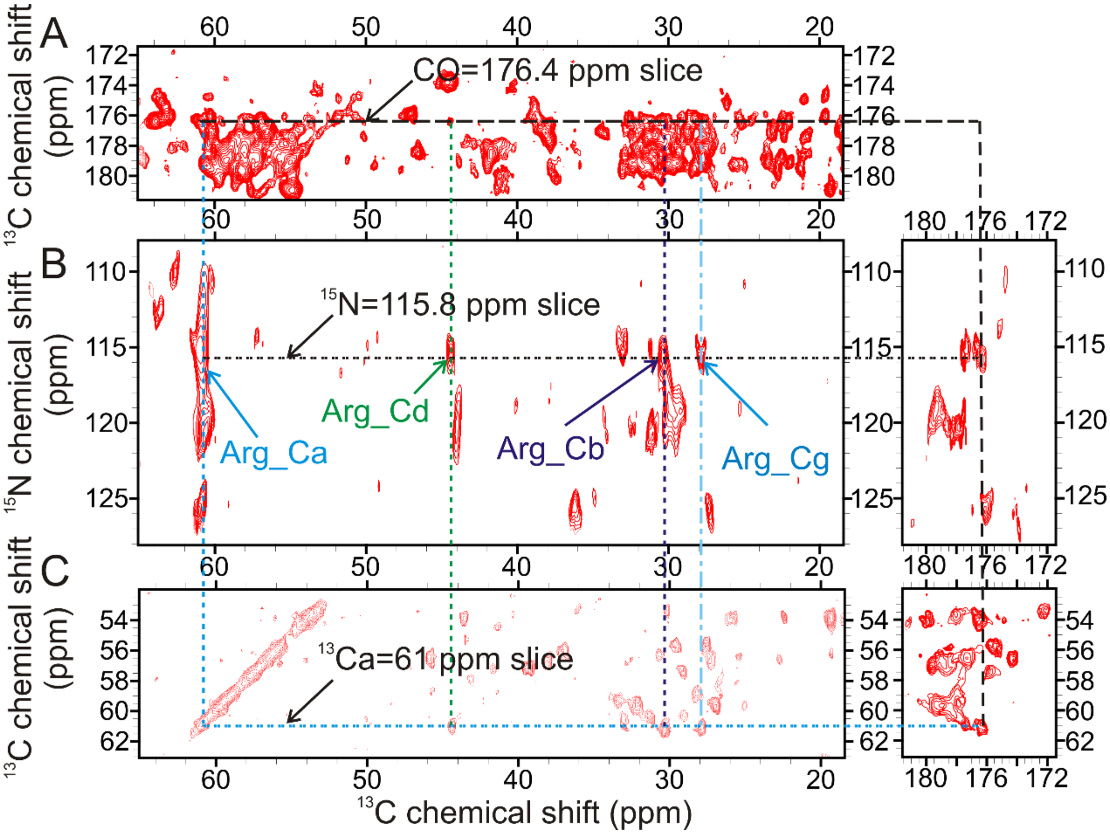


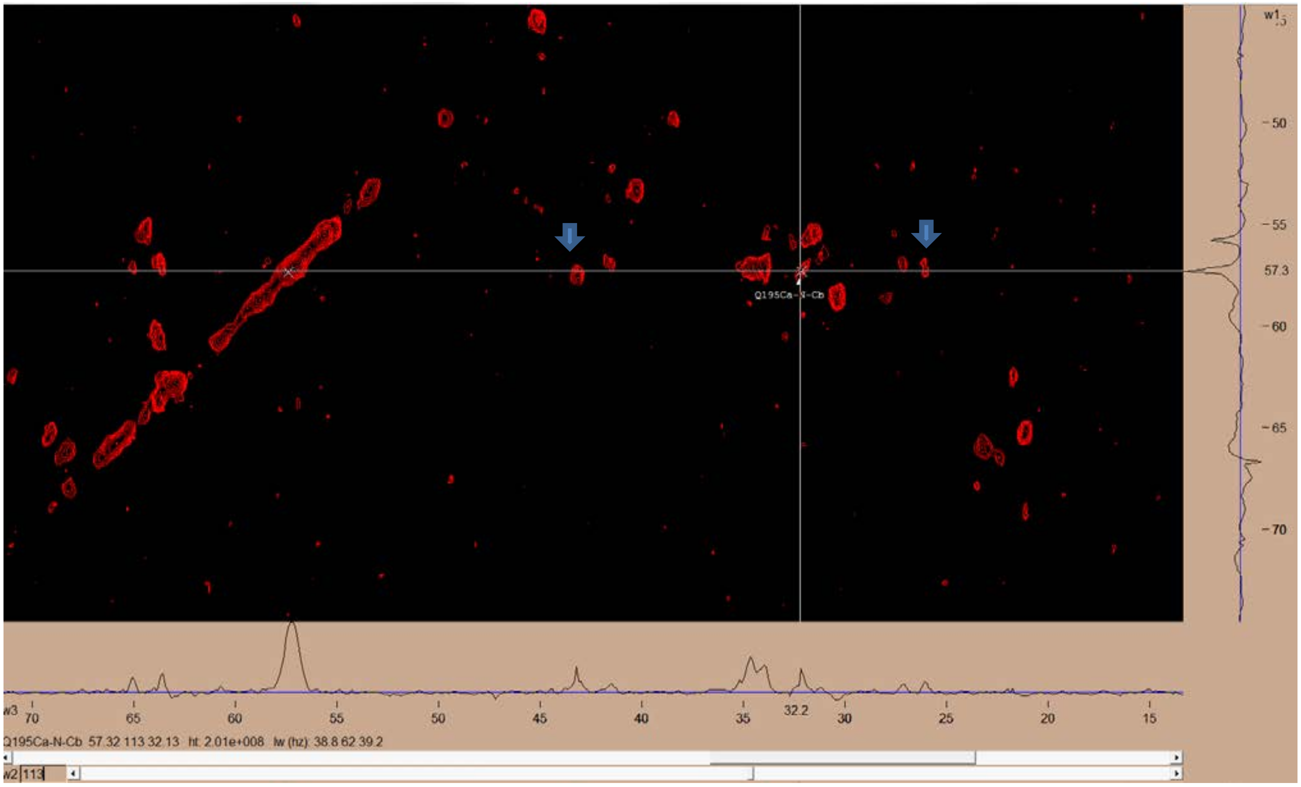
Figure 2. Assignments of congested NCOCX by exploiting well-resolved NCACX spectra. (A) The ^{15}N=119.1 pm plane of the NCOCX spectra is highly congested, with a large number of resonances at CO frequency (vertical axis) 176.4 ppm, indicated by the horizontal dashed line. (B) and (C) are the aliphatic and CO regions of the better resolved ^{13}C=61 ppm and ^{15}N=115.8 ppm planes of the NCACX spectra, respectively. They reveal intra-residue correlations between carbons and Ca of 61 ppm resonance in (B), or with the amide nitrogen of 115.8 ppm in (C). By inspection, only resonances at 44.5, 30.4 and 27.8 ppm are present in both NCACX and NCOCX spectra and correlated with CO=176.4 ppm and Ca=61 ppm. Therefore, they are assigned as an R according to the characteristic resonances of amino acids.

We also quantified the advantage of selective residue labeled sample in assignments. Distinction and unambiguous residue-type assignments (RSA) are the prerequisite for accurate sequential assignments. Many proteins contain multiple copies of the same residue, or residues with similar NMR characteristic resonance. Sparsely 13C labeled samples with glycerol expression can help RSA to a certain degree. But out of twenty residues, only ten residues exhibit definitive labeling patter by glycerol expression. For example, Arg, Lys, Gln and Leu exhibit similar resonance patterns, as shown in Table 1. There are a large number of them in RSV CA. Ambiguous assignments of these residues lead to limitations in sequential assignments.

Table 1. Characteristic resonance frequencies of Lys, Leu, Gln and Arg.

	Ca	Cb	Cg	Cd	Ce
Lys	54.2	32.6	24.6	29.1	41.9
Leu	53.1	41.7	27.1	25.1,23.3	
Gln	53.7	28.8	33.4	180.5	
Arg	54	30.2	26.8	43.3	159.6

If we don’t have the selective labeling, it will be hard to distinguish them. For example, Q 195



```
Q195Ca-N-Cb 57.32 113 32.13 ht 2.01e+008 lw (hz) 38.8 62 39.2
```

Figure 3. Assignments of Q195 would be impossible to distinguish from Arg in the spectrum of uniform ^{13}C and ^{15}N labeled sample. Only by comparing with the resonances in the spectrum of Arg selective 13C labeled sample can we exclude this possibility, shown in Figure 4.

As showing in Figure 3, you can see there are two more peaks in the same line with the marked, at (57.6, 112.3, 43.2) and (57.1,113.6, 26.08), indicated by blue arrows. These three peaks match with the typical chemical shifts of Arg listed in Table 1. But no corresponding resonances can be found in the spectrum of selective Arg labeled sample. Therefore, it affirmatively excluded the ambiguity to assign this residue as Arg.

In Figure 5, it is hard to assign the peak R141 to be arginine, if we do not have the R labeled spectrum.

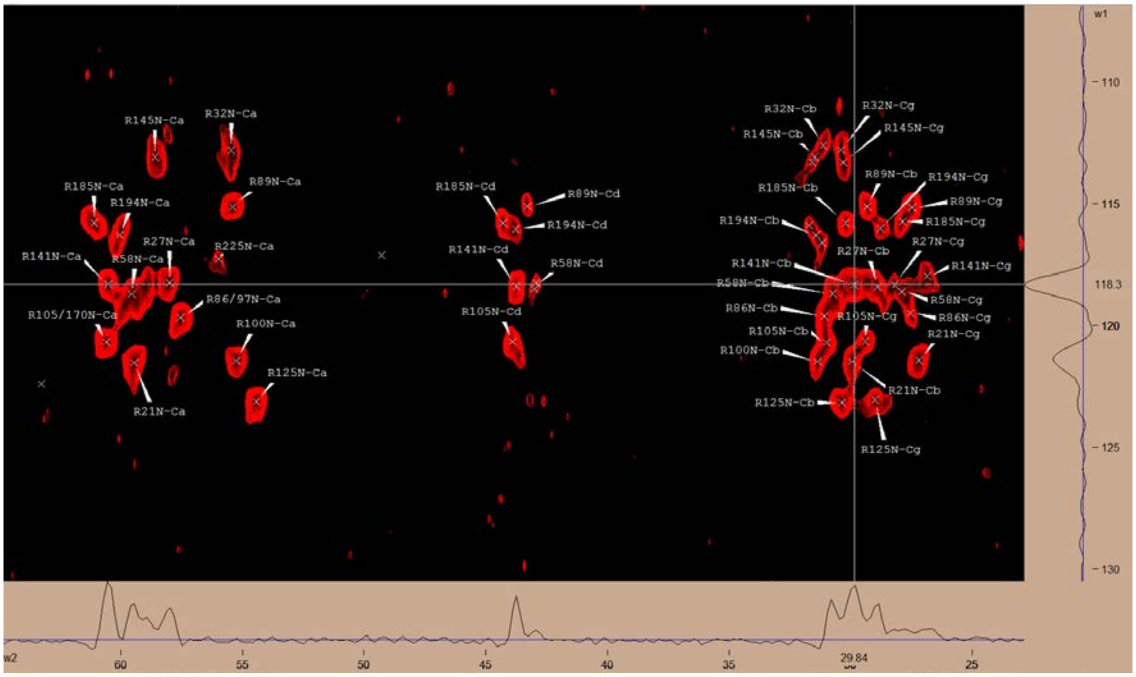


Figure 4. 2D NCACX spectrum of selective Arg labeled sample.

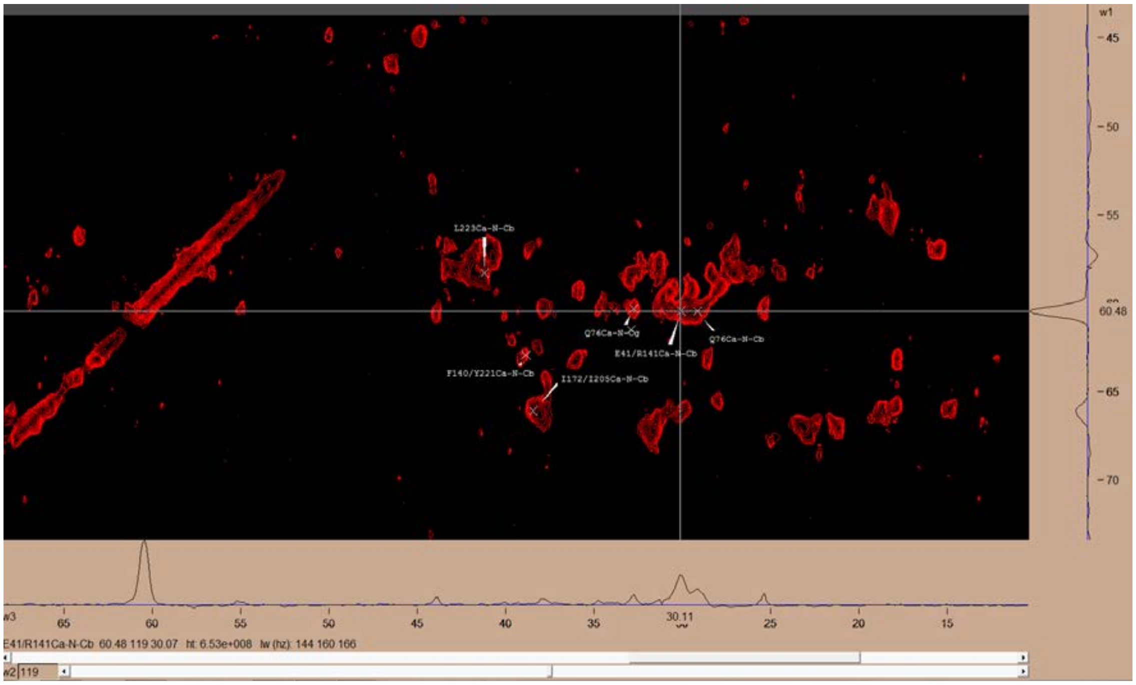


Figure 5. The peak at (119, 30.11, 60.48) could not be assigned as Arg 141 if were not with the Arg selective labeled samples.

But if we compare the R labeled spectrum, at N frequency 119ppm, Ca frequency 60.48 ppm, there is one peak. So we can assign this residue to Arg.

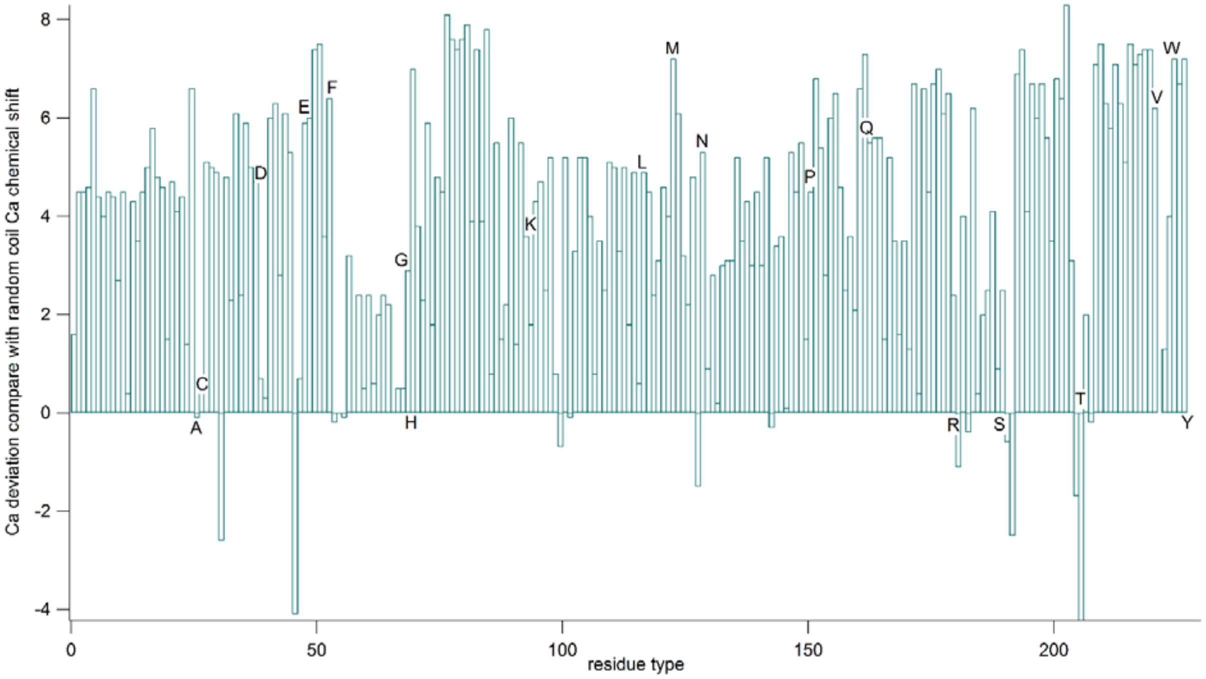
The number of sequential assigned residues was limited to 128, even with abovementioned assignment strategies. With two selective residues labeled samples(Arg and Leu), ambiguity of

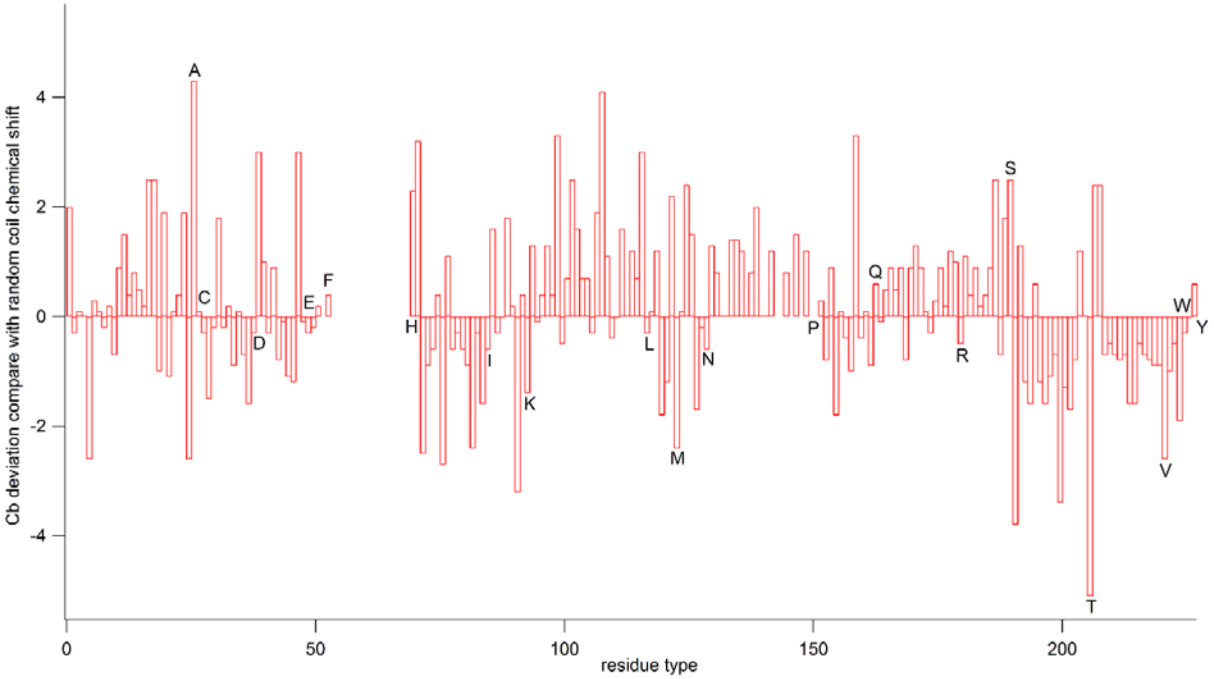
Arg, Lys, Pro and Leu was resolved and enhanced the number of sequential assigned residues to 234.

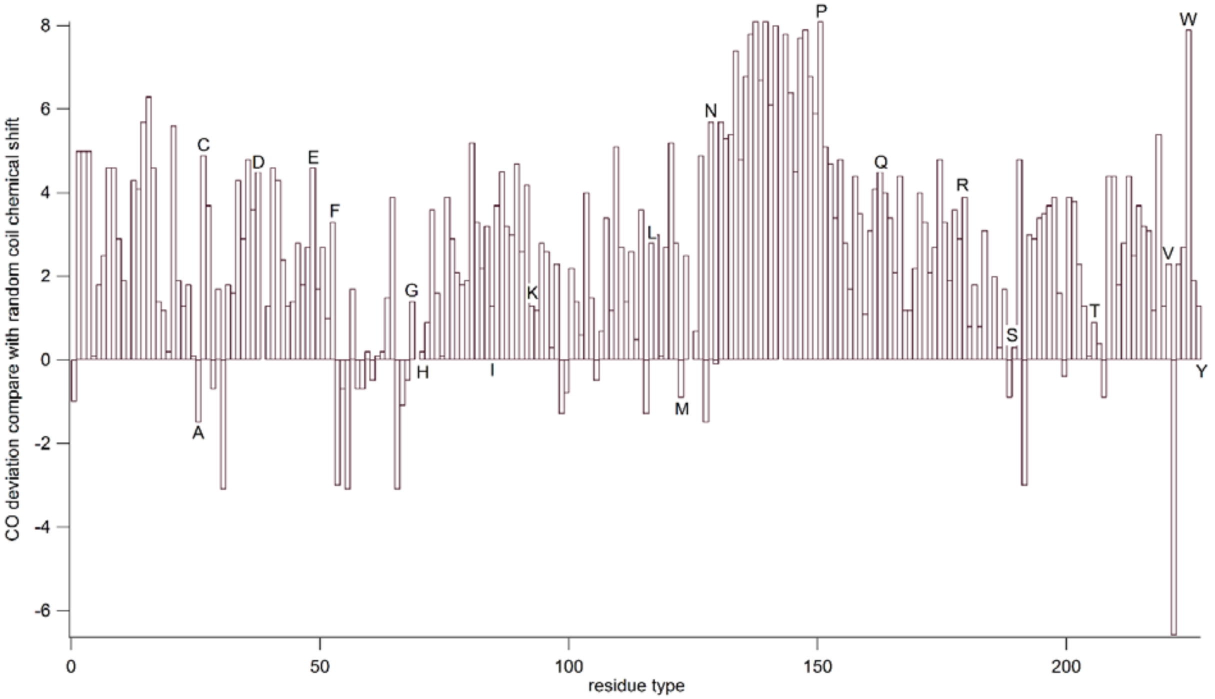
In addition to this technical development, we reviewed the details of the assignments and prepared a tutorial to help train NMR assignments.

Theoretically, the assignments of NMR resonances of protein spectra just need to consult the characteristic resonance patters of the 20 amino acids. However, due to the effect of local structure on CS, the experimental observed resonances may shift away from the ideal positions, and can become confusing for beginners. We collected the experimentally observed deviations from theoretical positions, as a summary tutorial to help beginners establish the correct expectation.

A. The assigned CS of carbons are compared to CS of residues in random coil conformation to derive the secondary structure. However, for beginners, it is not clear how large deviations should each site be allowed. After all, many residues possess similar characteristic resonance frequencies. If there is no clear range of CS variations of each amino acids, it is difficult for beginners to know how to accurately make RSA. Although Biological Magnetic Resonance Bank has a statistical distribution of the CS of each carbons in all assigned proteins available, their data do not differentiate proteins in different conformations and contains those with artificial amino acids, therefore, it is difficult to get a clear clue how to utilize the information. RSV CA comprises predominantly α-helices, and its assignment can be an ideal template for other proteins with α-helices. To provide a more effective guidance, we summarized the deviations of CS of each carbon sites as a guidance.







DISTRIBUTION A: Distribution approved for public release.

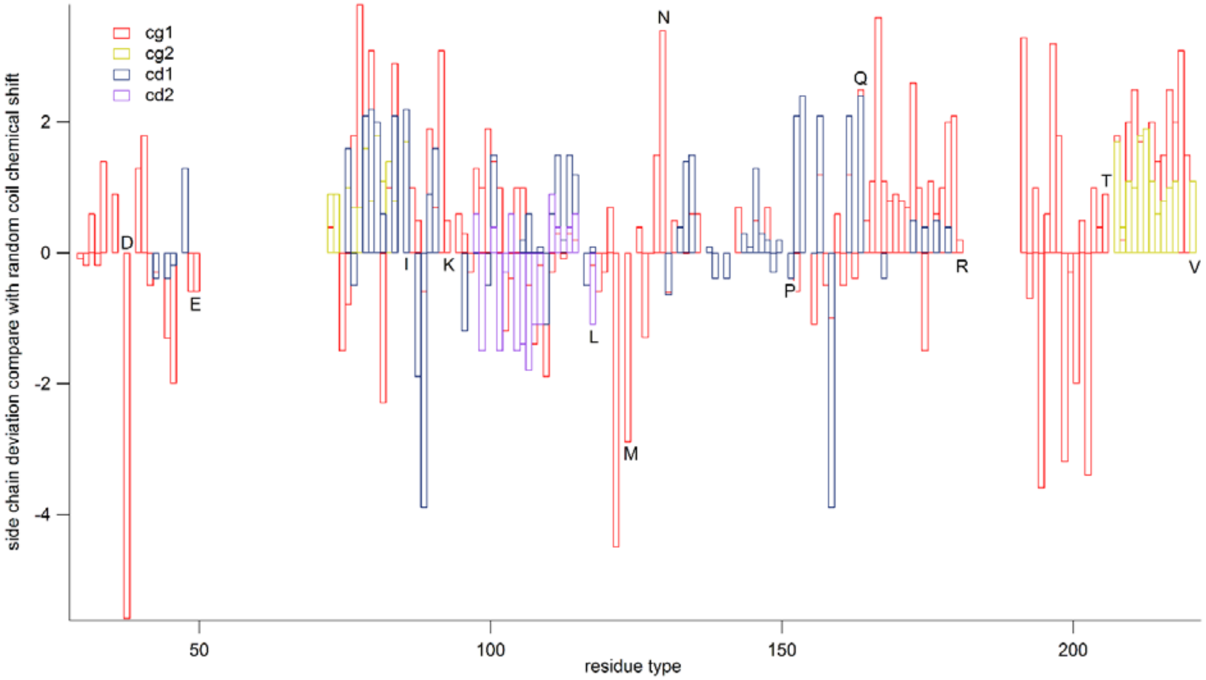


Figure 6. A-D indicate Ca, Cb, CO and sidechain deviations compare with random coil chemical shift, respectively. (A) shows Ca shifts up to 8ppm. (B) shows Cb shifts up to 4ppm. (C) shows CO shifts up to 8ppm. (D) shows sidechain shifts up to 4ppm. CO and Ca deviations are higher than Cb and sidechain for same residue. Different residues have different chemical shift deviation.

As Figure 6 shows, Ca and Co displays comparable deviations from random coil CS, normally larger than 5 ppm, to as large as 8 ppm. The deviations of side chain carbons are much smaller, usually smaller than 3 ppm. It shows the side-chain carbons are the best signature to determine the residue types.

B. Theoretically, all carbons within the same residue should exhibit identical ^{15}N and Ca frequency. However, in our NCACX spectra, we frequently encountered deviations from this ideal situation, where ^{15}N frequency for Ca, Cb, Co and side-chain could be slightly different(larger than 0.3 ppm, and in a few cases, up to 2 ppm. We confirmed they actually belong to the same residue by inspecting R and L labeled sample. Therefore, we choose the ^{15}N frequency of Cb as the resonance frequency of amide nitrogen for this residue. This kind of deviation from theoretical values could lead to mis-grouping resonances belonging to the same residues and request extra efforts of conformation.

Figure 7 demonstrates this deviation of resonance frequencies for atomic sites belonging to the same residue. Figure 4 shows the Arg selective labeled 2DNCACX spectrum (therefore, only Arg produce NMR signals in this spectrum). However, we can find not all of the peaks of Arg 194 are in the same line, which means there is deviation for their N frequencies. Most of the time, the N frequency of Ca consists with the N frequency of Cb, but starts from Cg, there is a little bit deviation. And also, the N frequency deviation of Cg is smaller than Cd, see the example of Arg58.

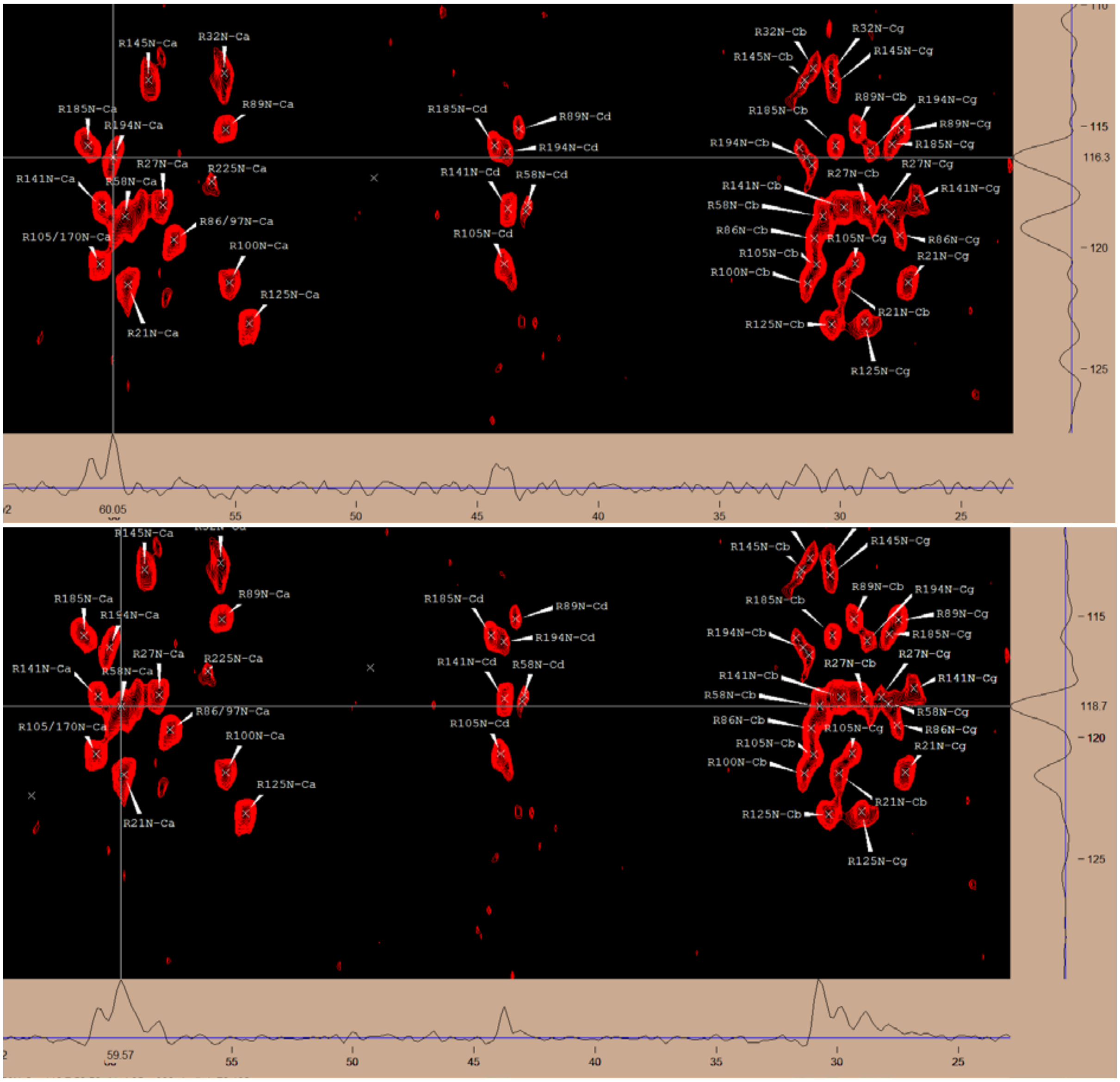


Figure 7. Deviations of resonances belonging to the same residue from ideal positions.

In the above figure, we can obviously find N frequency of Ca is 118.7 which is same as Cb, but the N frequencies of Cg and Cd are 118.6 and 118.5, separately. So the N frequency deviation of Cg is smaller than Cd.

(2). Establishment of the first atomic resolution structural model of RSV CA tubular assembly.

In the past six months, we started collaboration with Drs. Kingston and Mitra's group at University of Auckland (New Zealand). They provided 24 Å resolution cryo-EM image of RSV CA tube, which provides tertiary and quaternary constraints for RSV CA in the assembly. Combining with torsion angles derived from our ssNMR assignments, Dr. Fangqiang Zhu performed Molecular

Dynamics Flexible Fitting (MDFF) simulations and determined the first atomic resolution model of the RSV CA tubular assembly, as shown in Figure 8.

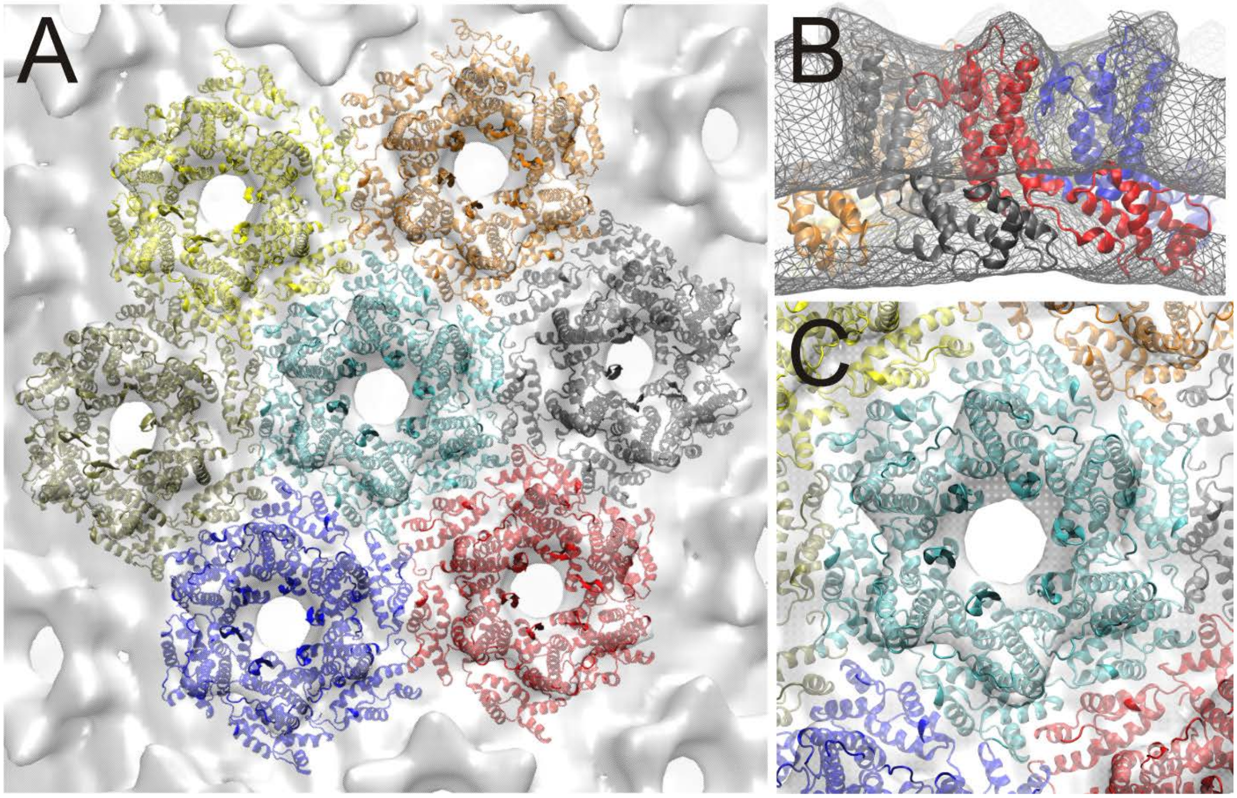


Figure 8. Atomic model obtained from MDFF displayed along with the cryoEM electron density map. (A) Seven hexamers on the tube surface. (B) Side view of a single hexamer, with each of the six monomers shown in a different color. (C) Top view of the central hexamer with its six neighbors partly shown.

Specifically, Our initial model was taken from an X-ray structure (PDB ID: 3TIR) of a planar assembly of RSV CAs [1]. The missing linker (residues 148-151) between the NTD and CTD and the truncated C-terminal tail (residues 227-237) in the crystal structure [1] were added back using the program MODELLER [2], thus resulting in a full-length CA monomer with the entire set of 237 residues. We then generated a CA hexamer according to the six-fold symmetry specified in the crystal structure [1].

In the next step, we performed a simulation on the CA hexamer generated above to convert the secondary structures according to the ssNMR data. The TalosN program [3] was used to predict the backbone torsions from the chemical shift, and assigned the φ and ψ angles for 231 of the 237 residues in a CA monomer. Among these assignments, 200 residues were evaluated as strong/generous by TalosN, whereas 31 residues were indicated as ambiguous or dynamic. In our simulation, we applied harmonic restraints on the backbone φ and ψ angles for all of the assigned residues. A spring constant of 50 kcal/mol/rad^2 was used for residues with the strong/generous status, and a weaker restraint of 25 kcal/mol/rad^2 was used for those with the ambiguous/dynamic status. The torsion assignment for residue E7 (with the warning status) was highly inconsistent with the local geometry in that region, and we thus did not apply torsion restraint on this residue. In addition, we applied torsion restraints (with a spring constant of 50 kcal/mol/rad^2) on all peptide bonds in the protein backbone to prevent them from being converted into the *cis* configuration [4] during the simulation. To maintain the conformation of the β-strand at the N-terminal, we also applied distance restraints (with a spring constant of 20 kcal/mol/Å^2) on four pairs of backbone H-bond donor/acceptor atoms in each monomer. The simulation was run with a dielectric constant of 80 and at a constant temperature of 600 K to allow sufficient flexibility in the secondary structures. In the meantime, each C_α atom was restrained (with a spring constant of 1 kcal/mol/Å^2) to its position in the crystal structure [1] to prevent large-scale conformational change of the protein in this stage. In addition, symmetry restraints (with a total spring constant of 200 kcal/mol/Å^2) were applied on the heavy atoms of the six monomers to maintain the six-fold symmetry as in the crystal structure [1]. The simulation for the single CA hexamer was run for 0.2 ns, followed by an energy minimization.

The structure obtained above is largely consistent with the backbone torsions predicted from ssNMR. We then docked this single hexamer to the electron density map from cryo-EM, using the Situs program [5]. The cryo-EM map shows the density of 784 CA hexamers in a tubular assembly, and exhibits a helical symmetry with a rotational angle of 192.67° and a rise of 5.85 Å. After the rigid-body docking of our hexamer structure to the cryo-EM map, we generated its six nearest neighbors according to the helical symmetry above, thus resulting in a patch of seven CA hexamers in the tubular assembly, with one hexamer surrounded by the other six.

In the next stage, we performed an MDFF simulation [6, 7] for the 7-hexamer system above. The MDFF technique establishes a grid-based external potential from the cryo-EM density map, such that protein atoms are attracted to the high-density regions of the map [6, 7]. In our simulation, all heavy atoms of the protein are subject to the MDFF potential with a scaling factor of 0.5. Similar to the single-hexamer simulation described earlier, this simulation was performed in vacuum with a dielectric constant of 80, and with the restraints on the backbone torsions (based on ssNMR), the peptide bonds, and the backbone H-bonds in the β-strand at the N-termini. However, the six-fold symmetry restraints on the monomers in a hexamer were not applied here. Instead, we took each hexamer as a single unit, and applied symmetry restraints [8] (with a total spring constant of 200 kcal/mol/Å^2) on the C_α atoms of the seven hexamer units in the system, based on the helical symmetry described earlier. In addition, to maintain the overall fold of protein domains, we applied domain restraints on the NTD hexamers and the individual CTDs. The domain restraints are in the form of harmonic potentials on the root-mean-square-deviation (RMSD) for the C_α atoms in the corresponding protein domains with respect to reference structures, with a total spring constant of 100 kcal/mol/Å^2. The six NTDs (residues 1-145) in the crystal structure [1] were treated as a single domain and used as the reference structure for the NTD hexamer restraints, and the CTD (residue 152-226) structure [1] was used as the reference for the domain restraints on the individual CTDs. The simulation was run for 7 ns, with the temperature initially at 600 K and gradually decreased to 300 K.

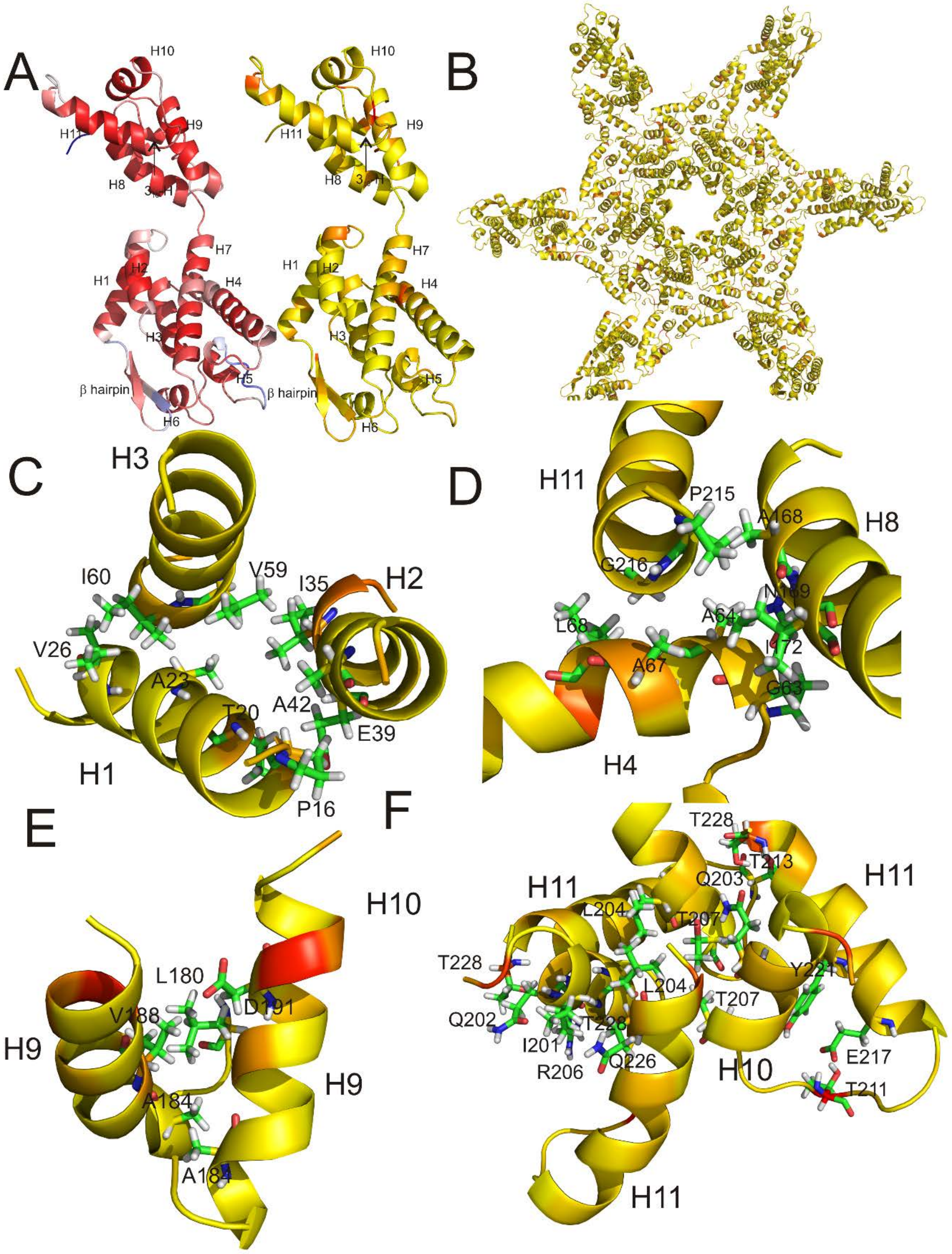


Figure 9. Structural model of the RSV CA tubular assembly. (A) Monomer of the RSV CA in the tubular assembly. The left is color coded according to the backbone dynamics (from red to blue for rigid to mobile),

and the right is color coded according to largest changes of the backbone (Ca, N, CO) ΔCS of each residue (from yellow to red for the least to the largest changes), indicating changes of backbone structure relative to that in the soluble state. (B) A RSV CA hexamer with adjacent subunits on the tube, color coded in the same manner as the left panel of (A). (C) to (F) are close-views of the NTD-NTD, NTD-CTD, dimer, and trimeric interfaces. Only those helices involved in close contacts are labeled. Backbones are color coded in the same manner as the left panel of (A). Residues inferred at close contacts at each interfaces are shown in stick mode.

Based on the 7-hexamer structure derived from the annealing simulation above, we further constructed a system with explicit water molecules and ions. The protein complex was solvated in a box of water molecules described by the TIP3P model [9], and ions were then added to represent a NaCl solution of 50 mM and to render the system electrically neutral. The resulting simulation system has a total of ~2.6 million atoms, and a size of ~239 Å × ~312 Å × ~348 Å.

We then performed a simulation for the explicit-solvent system above, with periodic boundary conditions, constant temperature (300 K) and constant pressure (1 atm) maintained by the Langevin piston method [10]. Full electrostatics was calculated every 4 fs using the particle-mesh Ewald method [11]. All restraints in the protein-only simulation described earlier were also applied in this explicit-solvent simulation. These include restraints on the helical symmetry, on H-bonds in the β-strands and on the backbone torsions, the grid potential based on the cryo-EM density map, and the domain restraints. The simulation was run for 5 ns, and was further continued by another 5-ns simulation in which the domain restraints were removed to allow more flexibility in the conformations of protein domains. Finally, we performed an energy minimization of the entire system using a stronger cryo-EM grid potential with a scaling factor of 1.

The final model shows four intermolecular interfaces stabilizing the tubular assembly, similar to those identified in RSV and HIV CA assemblies[1, 12-17], as shown in Figure 8. The NTD-NTD interface, involving the first three helices of CA, bundles NTDs together into individual hexamers, which are further stabilized by the NTD-CTD interface between helix H4 of each NTD, and helices H8 and H11 from the CTD of the neighboring subunit within the hexamer. The CTDs sit below the NTD-hexamers, and form dimeric (Helix H9) and trimeric (Helices H10 and H11) interactions with their immediate neighbors thereby creating the hexamer lattice.

Figure 9A shows a CA monomer in our final model. The RMSD between the final model with and without the Ca restraints from 3TIR is only 0.9 Å, suggesting that our modeling procedure has largely converged. The RSMD between a monomer defined in 3TIR and our model is 3.795 Å. It indicates that the obtained structural model is primarily guided by cryo-EM and ssNMR, rather than by the artificial restraints from 3TIR. Moreover, the average RMSD for the 42 monomers in the model is 2.69 Å, comparable to the 2.64 Å RMSD of the PDB entry 3J34 for the HIV CA tubular assembly[17] obtained with a cryo-EM map at much higher resolution.

Although no explicit constraints of contact interfaces were used in our MDFF modeling process, the final model shows that the RSV CA tube is stabilized by four intermolecular interfaces, similar to those identified previously in RSV and HIV CA assemblies[1, 12-17], as shown in Figure 9C to F: The NTD-NTD interface between the first three helices bundles NTDs together into individual hexamers, which is further stabilized by the NTD-CTD interface between H4 and 8. The CTDs form dimers across neighboring H9. They sit below NTD-hexamers, and connect neighboring NTD-hexamers into lattices, aided by the trimer interface between H10 and 11. Moreover, our final model enables residue-specific contacts, and the interacting residues at interfaces coincide with those exhibiting large change of torsion angles and ΔCS, which further lends support to our approach and model. To show this correlation, Figure 9 plots the backbone of subunits color coded according to the maximum of backbone ΔCS(Ca, CO, N). As shown in Figure 9C, the NTD-NTD interface is stabilized by contacts between T20 (ΔCS(N)=2.93 ppm) at H1 and E39 (ΔCS(N)=0.637 ppm) at H2, and between I35 (ΔCS(Ca)=-3.53 ppm) at H2 and V59 (ΔCS(N)=-0.801 ppm) at H3. Figure 9D shows the NTD-CTD interface: A64 (ΔCS(CO)=-1.94 ppm) at H4 is capped by N169 (ΔCS(CO)=1.017 ppm) at H8. The closest contacts at dimer interface are between D191 (ΔCS(CO)=0.574 ppm), V188 ((ΔCS(CO)=-3.34 ppm) and A184 (ΔCS(N)=-0.49 ppm) at neighboring H9, as shown in Figure 9E. At the trimer interface, K227 (ΔCS(Ca)=0.799 ppm, (ΔCS(Cb)=-1.5 ppm) at H11 is close to R225 (ΔCS(CO)=0.383 ppm) and Y221 (ΔCS(Ca)=1.703 ppm) of neighboring H11, and to T207 (ΔCS not available, it is A207 in solution NMR data in BMR_4384) of neighboring H10, as shown in Figure 9F. Moreover, some of their adjacent neighbors also show considerable ΔCS: At NTD-NTD interface, I60 shows ΔCS(N)=-2.222 ppm; at the dimer interface, C192 exhibits ΔCS(Ca)=-7.761 ppm, and F193 has a ΔCS(Ca)=3.995 ppm; at the trimer interface, Q226 displays ΔCS(Ca)=2.003 ppm and ΔCS(CO)=2.526 ppm, T228 also exhibits ΔCS(Ca)=-5.475 ppm and ΔCS(N)=-5.824 ppm.

Because the similar CS to the soluble RSV CA[18], a question naturally arises: if our ssNMR constraints actually impart any meaningful differences to the final model. To answer this question, we repeated the MDFF modeling simulation by replacing the torsion angles derived from ssNMR with soluble RSV CA. Overall, the final model obtained with soluble RSV CA NMR constraints does resemble the correct model, with 2.798 Å RMSD. But we find surprisingly large changes of intermolecular interfaces, as shown in Figure 10. Specifically, the crossing angles between H2 and 3 at the NTD-NTD interface changes from 154° in the correct model to 170°, besides the change of residues at this interface. Moreover, not all subunits in the final model form stable NTD-CTD contact, and when they do, the interface differs as well. The dimer interface displays the least difference, shown in Figure 10. At the CTD trimer interface, the original interlocking side-chains of R225 and K227 in neighboring subunits are disrupted, as shown in Figure 10. In addition, the residue-specific model enabled by our ssNMR constraints actually confirms the mutagenesis tests in recent report[19], as they form contacts critical to stability of the overall molecular structure. For example, close intramolecular contacts are observed between R145 and M37, between T207 and D179, between E41 and L137, between Q138 and Q76, between W153, V188, and L180. K173 interacts with P63 from neighboring subunits, belonging to the NTD-CTD interface. Collectively, our results suggest that, despite of largely conserved CS, the few localized regions with large CS change in RSV CA upon assembly do translate into important structural differences in 3D assembly model.

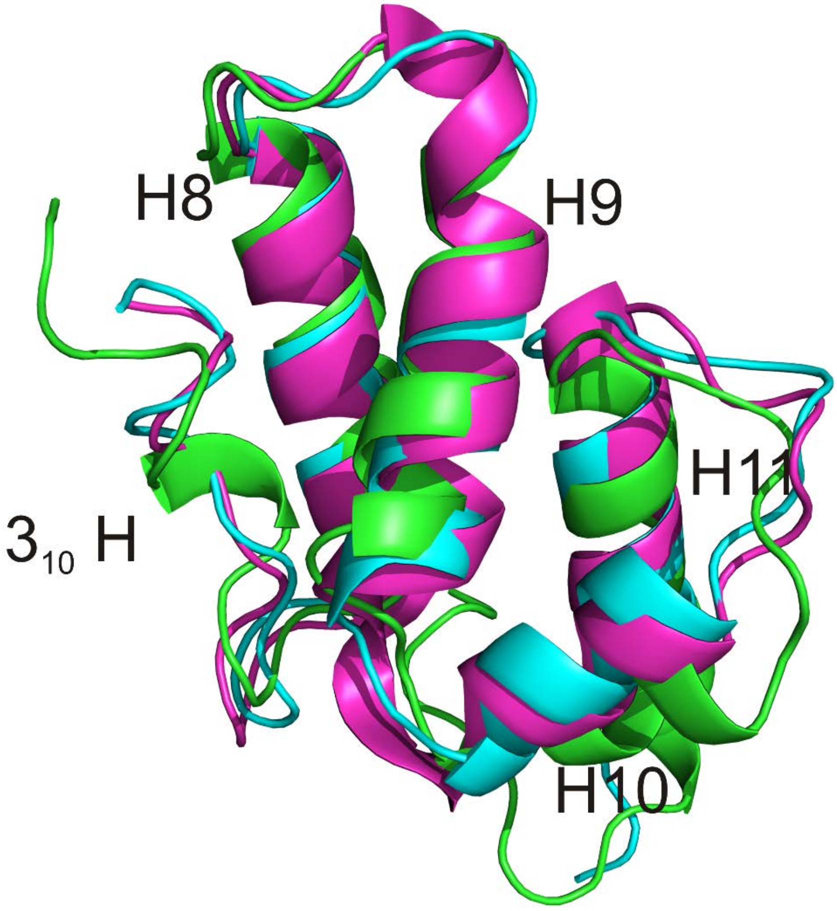


Figure 10. Conformational change of 3_{10} helix. Green: this work. Cyans: solution NMR RSV CA structure in PDB entry 1D1D.[18] Magentas: X-ray crystallography structure of flat hexameric sheet of RSV CA in PDB entry 3TIR.[1]

Although tertiary constraints within individual domains from 3TIR[1] were used at the initial stage of our modeling process, the tertiary and quaternary level structure of our model were later refined based entirely on the cryo-EM density map. This is demonstrated by the 3.795 Å RMSD and the different crossing angles between helices at interfaces between our model and those defined by 3TIR, summarized in Table 1. Moreover, the different crossing angles provide valuable insights to understand the formation of RSV CA tubes, as the flat sheets were proposed to be the precursory step to curl into tubular assembly[20-22]. Alternatively, if the tubes are assembled *de novo*, the different crossing angles implicate different assembly pathways, as shown by recent CG simulations[21, 22]. As shown in Table 1, among all interfaces, the NTD-NTD interface exhibits relatively small difference. Large change takes place at NTD-CTD and dimer interface. The crossing angle between H9 at dimer interface is quite similar to that formed at low pH (PDB entry 3G21)[23]. However, the closest contact is between H9 and 3_{10} H in 3G21, but in our model is between the neighboring H9. The 3_{10} H is displaced in our model further away from H9. The trimer interface also exhibits significant rearrangements. Although H10 and 11 form a similar contact angle, the structural rearrangements put neighboring H11 in close contacts in the tube while displacing neighboring H10, which is part of the trimer interface in the flat sheet.

Table 1. Crossing angles of helices at each interface. Angles were measured in pymol by anglebetweenhelices.py script. “NA” means that pair of helices are not in close contact at that interface. Standard deviations were calculated by averaging all subunits in the model.

Crossing angles between helices at interfaces	RSV CA tube, this work	RSV CA flat sheet in 3TIR.pdb	HIV CA tube in 3J34.pdb	HIV CA flat sheet in 4XFX.pdb
H1 &2 at NTD-NTD interface	146.2°±2.3°	158.9°±0°	152.6°±5.5°	149.5°±0°
H2 &3 at NTD-NTD interface	157.4°±1.7°	152.9°±0°	145.0°±6.6°	138.0°±0°
H4 &8 at NTD-CTD interface	85.7°±4.3°	35.6°±0°	74.8°±9.2°	71.5°±0°
H9 at dimer interface	61.2°±6.2°	52.9°±0°	68.6°±11.8°	66.2°±1.1°
H10 &11 at trimer interface	156.3°±7.3°	130.0°±26.9°*	148.8°±22.8°	148.7°±17.4°*
H10 &10 at trimer interface	65.8°±5.4°	52.3°±0°	39.9°±14.3°	62.3°±0°

* There are three pair of H10 and 11 at each trimer interface, they do exhibit different crossing angles, but the crossing angles between the same pair of H10 and 11 at different trimer interface in the crystal model are identical.

Combined, our analyses present new insights how RSV CA forms similar hexameric assemblies of different curvatures with largely conserved structure. Compared to the flat hexamer sheet[1], the large variations of helices crossing angles, which is a manifestation of structural rearrangements, are likely a consequence of the localized large change of torsion angles of the loops between helices, interdomain linker, and 3_{10} H. As is known, the 3_{10} H is the MHR, which intimately affects the stability and helices packing in CTD[24-30]. This conjecture is also consistent with the larger charges of crossing angles involves helices in CTDs. In addition, the CTD trimer and dimer interfaces in different subunits exhibit RMSD values of 2.40 and 2.08 Å, respectively, which is twice of the NTD-NTD interface (1.24 Å). In the flat hexamer model 3TIR, N169 was proposed to be the pivot to induce change of domain orientations that causes assembly polymorphism[1]. We do observe N169 form water mediated H-bonds at NTD-CTD interface, as shown in Figure 11B. However, the differences between 3TIR and our model cannot be reconciled by rotation around N169. Collectively, our model suggests that the polymorphism is caused by the flexibility associated with CTD, potentially resulted from the structural rearrangements in the interdomain linker, 3_{10} H and loops between helices in CTD.

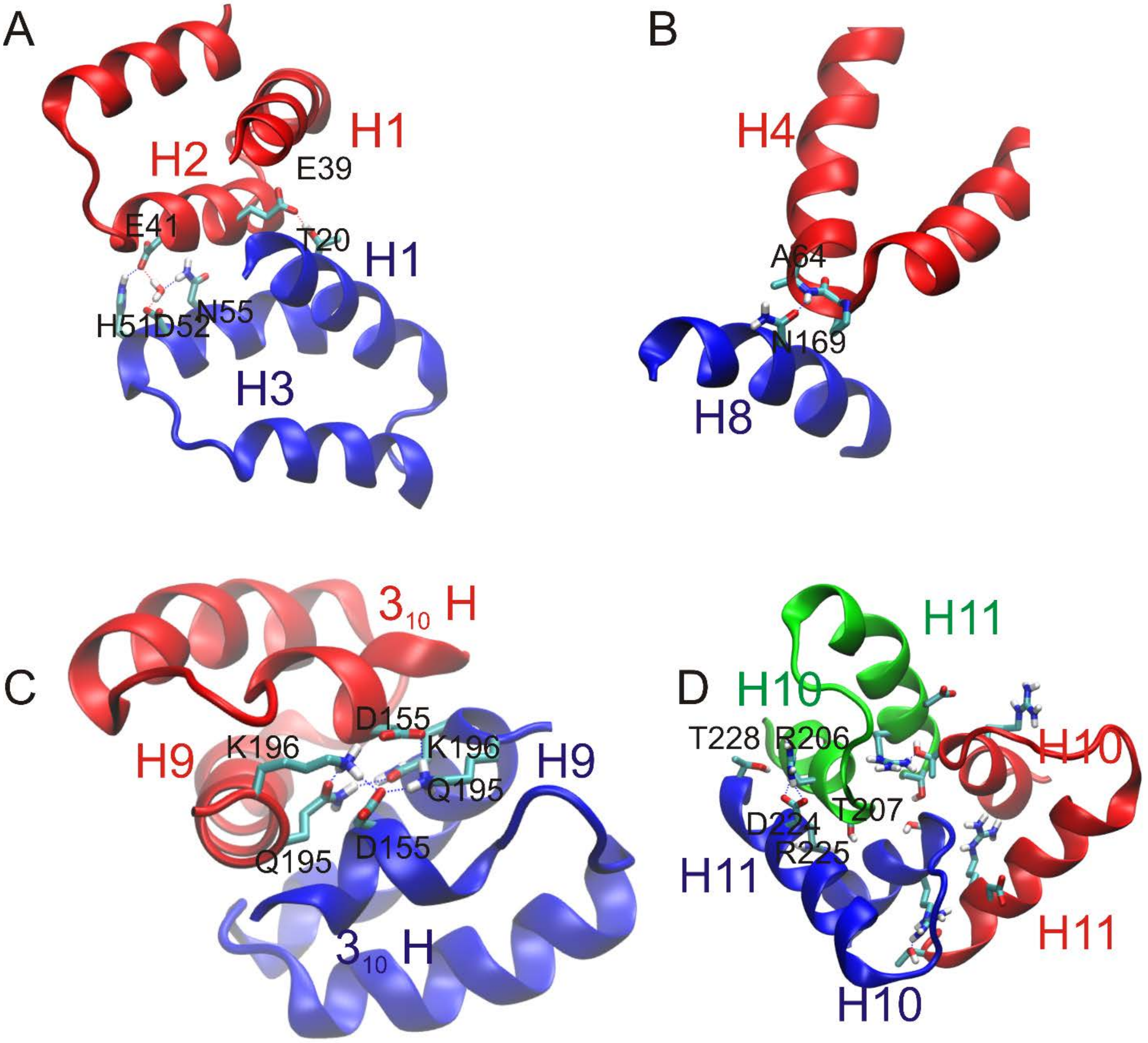


Figure 11. H-bonds identified at interfaces. (A) E39 H-bonds with K29 and T20 between H1 and H2, and E41 form H-bond with N55 at the NTD-NTD interface. (B) A64 form H-bond with N169 at the NTD-CTD interface. (C) H-bonds between D155, Q195 and K196 at the dimer interface. (D). H-bonds between R206, T207, R225, D224, and T228 at the trimer interface. Helices from difference subunits are in colored differently. Residues involving in H-bonds are represented in stick mode.

The HIV CA assembly is able to accommodate significant re-arrangements due to the existence of water-mediated H-bonds at interfaces [31]. Pervasive H-bonds are also observed at interfaces in our model, as shown in Figure 11. They contributed to the aforementioned malleable interfaces. As a comparison, we also applied similar analyses of crossing angles of helices at interfaces to HIV CA hexamer (PDB entry 4XFX) and tubular assemblies (PDB entry 3J34)[17, 31], shown in Table 1. Similar to RSV CA, HIV CA assemblies have well conserved NTD-NTD interface. On the other hand, the HIV CA trimer interface displays large change similar to RSV CA. But the dimer interface of HIV CA subunits is well conserved, in contrast. This is due to the strong dimer interface of HIV CA with a K_D~18 μM at neutral pH [32] while RSV CA stays as monomer at pH 5.7 and higher, and dimerizes only weakly at mildly acidic pH[13]. The NTD-CTD interface of HIV CA exhibits little change compared to that of RSV CA, indicating possibly a stronger interface. Compared to HIV, the NTD of RSV CA carries more positive charge[29], which may lead to a weaker and more flexible NTD-CTD interface. In summary, the change of helices crossing angles between RSV CA tube and flat sheet follows the order NTD-NTD<dimer<NTD-CTD~trimer, but they are ordered as dimer<NTD-CTD<NTD-NTD<trimer for HIV CA. This difference suggests that HIV and RSV CA tubular assembly may proceed along different pathways, as suggested by our recent CG simulations of HIV CA assemblies[21, 22].

(3). Progress on sequential assignments of spherical assembly of RSV CA.

Meanwhile, we continue our characterization of RSV CA spherical assembly (Figure 12). It turns out the spherical assembly is much tougher to work with. The sample seems to be very easy to deteriorate. Every sample we prepared lost its resolution after one spectra acquisition. It takes 6 weeks or more to prepare one sample. The work load and cost quickly add up.

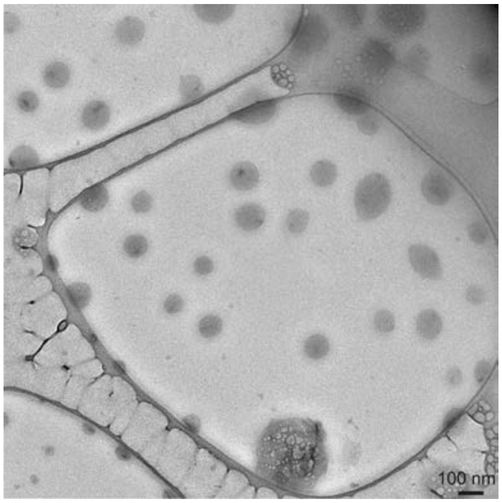


Figure 11. Negatively stained TEM image of spherical assembly of RSV.

So far, we have acquired 2D and 3D NCACX and NCOCX. The signal to noise (s/n) of these spectra are not optimal. More spectra will be acquired to enhance the s/n. Nonetheless, we have assigned 110 residues out of the 237-residue RSV CA in spherical assembly, aided with previous assignments of the RSV CA tubular assembly. Part of the assignments are labeled and shown in Figure 12, in the 2D NCACX spectra. Given time, the rest will be assigned, and atomic model of the spherical assembly will be established in collaboration with Drs. Kingston and Mitra at University of Auckland.

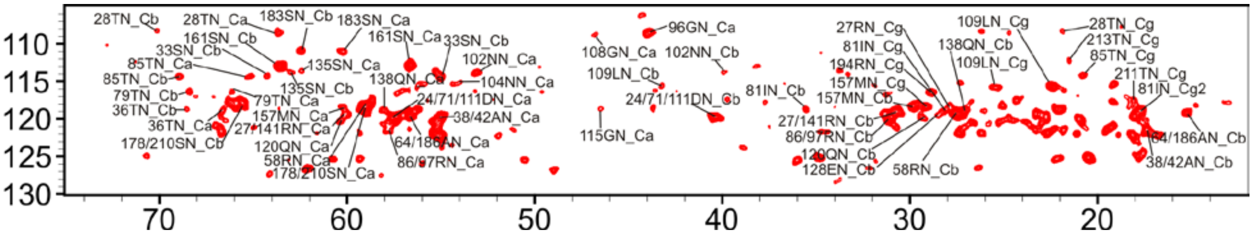


Figure 12. Partial assignments of the RSV CA in spherical assembly.

D. Supported personnel

In this half a year, the following personnel was being supported by this grant:
Tommy Boykin (graduate student) was partially supported by the grant.

E. Collaborations

We owe our gratitude to our wonderful collaborators, and they include:
Drs. Zhehong Gan and Ivan Hung at the National High Magnetic Field Laboratory (NHMF). They helped us acquired NMR spectra at 900 MHz field.
Dr. Fangqiang Zhu at IUPUI. He did most of the work to establish the residue-specific interaction potential.
Drs. Klaus Schulten and Juan Perilla at UIUC. They provided us the 303 ns all atom MD simulation.
Dr. Richard Kingston, Alok K. Mitra, and Ambroise Desfosses at University of Auckland, New Zealand. They provided Cryo-Electron Microscopy image to combine our NMR constraints so that we established the first atomic resolution model of the tubular assembly of RSV CA.
Dr. Rebecca Craven at Penn State contributed to our manuscript preparation, as well as the plasmid of RSV CA.

F. Publications

In this final half a year period, we have two publications:

1. **Bo Chen**. “HIV capsid assembly, mechanism and structure.”, invited (2016), 55(18), 2539-2552, Current Topics, ACS Biochemistry.
2. Jaekyun Jeon, Xin Qiao, Ivan Hung, Alok K. Mitra, Ambroise Desfosses, Daniel Huang, Peter L. Gor’kov, Rebecca C. Craven, Richard L. Kingston, Zhehong Gan, Fangqiang Zhu, and **Bo Chen**. “Structural model of the tubular assembly of the Rous sarcoma virus capsid protein.”, Submitted to JACS on Nov. 18.

G. Interactions/Transitions

Dr. Chen established collaboration with Drs. Richard L. Kingston and Alok K. Mitra at University of Auckland.

Xin Qiao, Dr. Chen’s student, presented the ssNMR assignment strategy as a poster presentation titled “Methods to assign 3D ssNMR spectra according to the assignment of RSV capsid protein” at Southeastern Magnetic Resonance Conferene (Oct 14-16, Atlanta, GA) and won Travel Award.

Dr. Chen was invited to be the biochemistry session chair for 2016 American Chemical Society Florida Section (May 5-7, 2016) and presented invited talk “Mechanism of the polymorphism and curvature control of the HIV capsid protein assemblies probed by a novel coarse grain model”.

Dr. Chen was invited to give an invited talk on March 15th, 2017, at 2017 American Physical Society Annual March Meeting, titled “Structural characterization of the Rous Sarcoma Virus capsid protein in its tubular assembly and simulations of the self-assemblies of the HIV capsid protein”

References:

[1] G.D. Bailey, J.-K. Hyun, A.K. Mitra, R.L. Kingston, A Structural Model for the Generation of Continuous Curvature on the Surface of a Retroviral Capsid, J. Mol. Biol., 417 (2012) 212-223.
[2] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol., 234 (1993) 779-815.

[3] Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks, Journal of biomolecular NMR, 56 (2013) 227-241.
[4] E. Schreiner, L.G. Trabuco, P.L. Freddolino, K. Schulten, Stereochemical errors and their implications for molecular dynamics simulations, BMC bioinformatics, 12 (2011) 190.
[5] W. Wriggers, Conventions and workflows for using Situs, Acta Crystallogr. D. Biol. Crystallogr., 68 (2012) 344-351.
[6] L.G. Trabuco, E. Villa, K. Mitra, J. Frank, K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics, Structure, 16 (2008) 673-683.
[7] L.G. Trabuco, E. Villa, E. Schreiner, C.B. Harrison, K. Schulten, Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography, Methods, 49 (2009) 174-180.
[8] K.Y. Chan, J. Gumbart, R. McGreevy, J.M. Watermeyer, B.T. Sewell, K. Schulten, Symmetry-restrained flexible fitting for symmetric EM maps, Structure, 19 (2011) 1211-1218.
[9] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, J. Chem. Phys., 79 (1983) 926-935.
[10] S.E. Feller, Y.H. Zhang, R.W. Pastor, B.R. Brooks, Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method, J. Chem. Phys., 103 (1995) 4613-4621.
[11] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, J. Chem. Phys., 98 (1993) 10089-10092.
[12] G. Cardone, J.G. Purdy, N. Cheng, R.C. Craven, A.C. Steven, Visualization of a missing link in retrovirus capsid assembly, Nature, 457 (2009) 694-U693.
[13] J.-K. Hyun, M. Radjainia, R.L. Kingston, A.K. Mitra, Proton-driven Assembly of the Rous Sarcoma Virus Capsid Protein Results in the Formation of Icosahedral Particles, J. Biol. Chem., 285 (2010) 15056-15064.
[14] O. Pornillos, B.K. Ganser-Pornillos, B.N. Kelly, Y.Z. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, M. Yeager, X-Ray Structures of the Hexameric Building Block of the HIV Capsid, Cell, 137 (2009) 1282-1292.
[15] O. Pornillos, B.K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid, Nature, 469 (2011) 424-+.
[16] I.-J.L. Byeon, X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, A.M. Gronenborn, Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function, Cell, 139 (2009) 780-790.
[17] G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics, Nature, 497 (2013) 643-646.
[18] R. Campos-Olivas, J.L. Newman, M.F. Summers, Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses, J. Mol. Biol., 296 (2000) 633-649.
[19] B.C.G. Katrina J. Heyrana, Juan R Perilla, Tam-Linh N. Nguyen, Matthew R. England, Maria C. Bewley, Klaus Schulten, and Rebecca C. Craven, Contributions of charged residues in structurally dynamic capsid surface loops to Rous sarcoma virus assembly, J. Virol., 90 (2016).
[20] Z. Yu, M.J. Dobro, C.L. Woodward, A. Levandovsky, C.M. Danielson, V. Sandrin, J. Shi, C. Aiken, R. Zandi, T.J. Hope, G.J. Jensen, Unclosed HIV-1 Capsids Suggest a Curled Sheet Model of Assembly, J. Mol. Biol., 425 (2013) 112-123.
[21] J.J. Xin Qiao, Jeff Weber, Fangqiang Zhu, and Bo Chen, Mechanism of polymorphism and curvature of HIV capsid assemblies probed by 3D simulations with a novel coarse grain model, BBA - General Subjects, 1850 (2015) 2353-2367.
[22] J.J. Xin Qiao, Jeff Weber, Fangqiang Zhu, and Bo Chen, Construction of a novel coarse grain model for simulations of HIV capsid assembly to capture the backbone structure and inter-domain motions in solution, Data in Brief, 5 (2015) 506-512.

[23] G.D. Bailey, J.K. Hyun, A.K. Mitra, R.L. Kingston, Proton-Linked Dimerization of a Retroviral Capsid Protein Initiates Capsid Assembly, Structure, 17 (2009) 737-748.

[24] J.B. Bowzard, J.W. Wills, R.C. Craven, Second-site suppressors of Rous sarcoma virus CA mutations: Evidence for interdomain interactions, J. Virol., 75 (2001) 6850-6856.

[25] C. Butan, P.M. Lokhandwala, J.G. Purdy, G. Cardone, R.C. Craven, A.C. Steven, Suppression of a Morphogenic Mutant in Rous Sarcoma Virus Capsid Protein by a Second-Site Mutation: a Cryoelectron Tomography Study, J. Virol., 84 (2010) 6377-6386.

[26] R.C. Craven, A.E. Leuredupree, R.A. Weldon, J.W. Wills, GENETIC-ANALYSIS OF THE MAJOR HOMOLOGY REGION OF THE ROUS-SARCOMA VIRUS GAG PROTEIN, J. Virol., 69 (1995) 4213-4227.

[27] P.M. Dalessio, R.C. Craven, P.M. Lokhandwala, I.J. Ropson, Lethal mutations in the major homology region and their suppressors act by modulating the dimerization of the rous sarcoma virus capsid protein C-terminal domain, Proteins-Structure Function and Bioinformatics, 81 (2013) 316-325.

[28] P.M. Lokhandwala, T.-L.N. Nguyen, J.B. Bowzard, R.C. Craven, Cooperative role of the MHR and the CA dimerization helix in the maturation of the functional retrovirus capsid, Virology, 376 (2008) 191-198.

[29] J.G. Purdy, J.M. Flanagan, I.J. Ropson, R.C. Craven, Retroviral Capsid Assembly: A Role for the CA Dimer in Initiation, J. Mol. Biol., 389 (2009) 438-451.

[30] J.G. Purdy, J.M. Flanagan, I.J. Ropson, K.E. Rennoll-Bankert, R.C. Craven, Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly, J. Virol., 82 (2008) 5951-5961.

[31] A.T. Gres, K.A. Kirby, V.N. KewalRamani, J.J. Tanner, O. Pornillos, S.G. Sarafianos, X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability, Science, 349 (2015) 99-103.

[32] T.R. Gamble, S.H. Yoo, F.F. Vajdos, U.K. vonSchwedler, D.K. Worthylake, H. Wang, J.P. McCutcheon, W.I. Sundquist, C.P. Hill, Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein, Science, 278 (1997) 849-853.